

Automated Classification and Scoring of Smooth Pursuit Eye Movements in Presence of Fixations and Saccades

Alex Karpov
Department of Computer Science
Texas State University-San Marcos
ak26@txstate.edu

Oleg Komogortsev
Department of Computer Science
Texas State University-San Marcos
ok11@txstate.edu

Abstract

Ternary eye movement classification, which separates fixations, saccades, and smooth pursuit from the raw eye positional data, is extremely challenging. This paper develops new and modifies existing eye tracking algorithms to perform meaningful ternary classification. To aid the purpose of meaningful automated ternary classification a set of qualitative and quantitative scores is introduced to facilitate the assessment of classification performance. Experimental evaluation of the proposed methods is conducted using eye movement records recorded from 11 subjects at 1000Hz in response to a step and ramp stimuli eliciting fixations, saccades, and smooth pursuit. Performance results for the automated and manual classification are reported.

CR Categories: I.6.4 [Simulation and Modeling]: Model Validation and Analysis; J.7 [Computers in Other Systems]: Process control, Real time.

Keywords: Eye movements, classification, algorithm, analysis, scoring, metrics, smooth-pursuit.

1 Introduction

The identification of the basic eye movement types from a noisy and frequently inaccurate raw eye positional signal is of utmost importance to the researchers and practitioners that employ eye trackers in their studies. Human Visual System (HVS) primarily exhibits six eye movement types: fixations, saccades, smooth pursuits, optokinetic reflex, vestibulo-ocular reflex, and vergence [1]. Among those eye movement types fixations, saccades, and smooth pursuit are most frequently studied. The following brief definitions can be provided for these eye movement types: eye fixation is an eye movement that keeps an eye gaze stable on selected stationary target, saccade is a very rapid eye rotation moving the eye from one fixation point to the next, smooth pursuit (SP) is eye movement that follows a moving object with a purpose of keeping the object on a high acuity vision zone called the fovea [1]. Eye fixations are frequently employed for human computer interaction as an input modality [2], saccades and smooth pursuits are frequently employed to diagnose pathologies of the HVS or assessing HVS performance in clinical populations [3].

Accurate automated eye movement classification is exceedingly difficult due to the noise and inaccuracies inherited from the eye

tracking equipment, dynamics of the HVS behavior, and variability between- and within- eye movement classification algorithms. Variation of single threshold value, in cases when only fixations and saccades are classified, is reported to substantially affect metrics such as number of detected saccades and fixations, average fixation duration and saccade amplitude [4-6]. Frequently researchers perform manual classification to avoid misidentification issues associated with automated algorithms. However in such cases classification becomes a very long and tedious process. Selection of the thresholds that provide meaningful classification is frequently done empirically with default values suggested by either eye tracking vendors or related literature. Given rapid developments of the eye tracking technologies that vary in hardware, sampling frequencies, and calibration algorithms [7] it is easy to “copy and paste” suggested thresholds; however it is hard to validate classification accuracy. During empirical threshold selection by “eye balling” small part of the classified data it is easy to misclassify some of recordings or misidentify corrective behavior such as corrected undershoots, overshoots, dynamic saccades etc. [1].

It is hard to define meaningfulness of the automated classification given a threshold value. For example it is possible to assume that quality of saccade detection can be ultimately judged by such properties as amplitude-duration relationship, main-sequence relationship, and saccades’ waveform [1]. However such characteristics are well established only for horizontal rightward saccades starting from the primary eye position [8] and are not well researched for other saccade groups. Recently Komogortsev and colleagues have proposed a set of behavioral scores with a purpose of selecting a meaningful classification threshold using a fixed stimulus [5]. Behavioral scores assume that amount of saccadic and fixational behavior encoded in a simple step-stimulus is matched by the HVS of a normal person therefore providing an opportunity to find a threshold value that ensures such performance. Researchers reported that thresholds selected according to these criteria provided meaningful classification results [5].

The goal of automated threshold selection for a fixed stimulus with subsequent employment of the same threshold for dynamic stimuli is particularly attractive because fixed stimulus is already presented as a part of the calibration procedure. Recording equipment’s performance for a given setup and subject is unlikely to change from calibration to the actual recording. Therefore, it is possible to assume that selected thresholds would continue to provide meaningful classification performance even during presentation of the stimuli that is different from the calibration.

Automated classification of SP in the presence of fixations and saccades is even more difficult task and continues to be a topic of an active research [9, 10]. Most difficult part of ternary eye movement classification is separation between fixations and SP. Two main factors contribute to the challenge: a) a fixation consists of the three sub-movement types such as tremor, drift, and

microsaccades. As a result a velocity range during a fixation (velocities up to 30°/s are possible during micro-saccades as computed by the main sequence relationship [1]) and SP (velocity up to 90°/s is reported in [11]) overlap. b) eye tracking noise further blurs quantitative boundaries between fixation and pursuit.

Given the importance of ternary eye movement classification and its challenges it is necessary to find out what degree of meaningful classification can be obtained and if such meaningful classification performance can be verified by a set of objective behavioral scores.

To start answering these questions this work: 1) introduces behavior scores related to SP, 2) proposes an algorithm for ternary eye movement classification 3) evaluates automated and manual ternary classification based on the proposed scores.

2 Overview

2.1 Classification of Fixations and Saccades

In general eye movement classification algorithms consider different properties of the signal that is captured by an eye tracker. In case when fixations have to be separated from saccades classification algorithms can be broken into the following groups: 1) position-based – Dispersion Threshold Identification (I-DT), Minimum Spanning Tree Identification (I-MST), 2) velocity-based - Velocity Threshold Identification (I-VT), Hidden Markov Model Identification (I-HMM), Kalman Filter Identification (I-KF), 3) acceleration-based – Finite Input Response Filter Identification (I-FIR) [5, 12, 13]. To the best of our knowledge these algorithms have not been “successfully” applied to the ternary classification.

2.2 Human Visual System Performance during Pursuit Stimuli

An SP is done via three phases [14]: 1) SP latency when the brain programs the movement for approximately 120ms, 2) corrective saccade that brings the target to the fovea, 3) a final SP movement.

Considering the description above it might appear that derivation of a SP classification algorithm is relatively simple. Such intuition would be based on a notion that SP is something average between very fast saccades and stationary fixations. Subsequently, it is possible to think that the use of a second threshold in I-VT algorithm to separate sandwiched SP signal between the fixations and saccades would be enough to ensure meaningful SP classification. However, there are several factors that prevent this simple thresholding scheme from being successful.

The first factor is jitter during fixations. Jitter is frequently caused by the inaccuracies in the eye tracker’s gaze position estimation. Good accuracy performance is varied in the range of 0.25°-1°. The second factor is presence of miniature eye movements such as drift, micro-saccades, and tremor [15] which all contribute to the jitter. Such conditions provide a high spread of the amplitudes for the positional (e.g., up to 1.5°) and velocity (e.g., up to 40°/s) signal. This spread does not greatly impact classification accuracy if only fixations and saccade are present, however in cases of SP of low velocity (e.g. 20-40°/s) the results of classification might be poor. The third factor is variation of the SP behavior between people. After beginning of the target’s motion some delay in the HVS response may occur. For example SP latency (the time dif-

ference between the onset of SP stimuli and actual SP movement that defines programming phase) can reach up to tens of milliseconds [16-19] and is dependent on such factors as fatigue or/and lack of concentration. To match target’s and eyes’ locations in case of a lagging performance HVS exhibits series of corrective saccades interspersed by the short periods of SP [1]. The frequencies and amplitudes of corrective saccades are subject dependent [20]. In general humans are capable of matching velocities of up to 90°/s during SP [11], however for some people saturation limit occurs for velocities of 30-40°/s with higher velocities being compensated by a series of corrective saccades and fixations [18].

2.3 Previous Algorithms for Automated Classification of Smooth Pursuit

In the previous research the separation of SP was done in cases when only SP and saccades were present. For example a single threshold-based algorithm was employed by [21]. The researchers used a velocity threshold of 50°/s to separate saccades from SP. All sequences of samples with velocity greater than threshold were checked upon matching main-sequence relationships. If a sequence of points met these criteria than all its samples were marked as a saccade. Otherwise the samples were discarded. Bahill was able to use main sequence relationship as a criterion for meaningful classification because only horizontal saccades were considered.

For ternary classification an interesting approach was proposed by San Agustin [9] and further enhanced by Larsson [10]. The approach monitors the direction of movement and rate of movement to separate fixations from SP. More details are provided in Section 4.

3 Behavior Scores for SP Classification

Considering the multitude of factors affecting SP performance and especially between-subject variability, it is important to develop simple metrics that can assess automated eye movement classification performance signaling the cases of classification success or failures with an ultimate goal of suggesting parameters/thresholds for meaningful classification.

Previously Komogortsev and colleagues [5] created a set of behavior scores that allowed assessment of classification quality or even determining the optimal threshold values when only fixations and saccades were present. This work continues in the same direction fine-tuning already existing scores and creating additional scores to assess meaningfulness of ternary classification. For the purposes of the initial assessment behavioral scores assume that the amount of fixational, saccadic, and SP behavior encoded in step and ramp stimuli is matched by the HVS in a normal subject.

3.1 Scores for fixation and saccades only

Komogortsev et al. [5] originally proposed three behavior scores namely Fixation Quantitative Score (FQnS), Fixation Qualitative Score (FQIS), and Saccade Quantitative Score (SQnS). The scores were originally designed to measure classification quality if only fixations and saccades are present in the raw eye positional trace. We perform following additions and modifications that allow extending the utility of behavior scores for ternary classification.

| Onset time, ms | Length, ms | Stimulus onset coordinates | | Stimulus Signal | |
|----------------|------------|----------------------------|--------|-----------------|---------|
| | | X | Y | A | V |
| 1000 | 709 | 0 | 0.00 | 14.2° | 20.1°/s |
| 2709 | 606 | -10 | 10.15 | 14.2° | 23.5°/s |
| 4314 | - | 0 | 0.00 | 14.2° | - |
| 5314 | - | -10 | 10.15 | 14.2° | - |
| 6315 | 531 | 0 | 0.00 | 14.2° | 26.8°/s |
| 7846 | 471 | -10 | -10.16 | 14.2° | 30.2°/s |
| 9316 | - | 0 | 0.00 | 14.2° | - |
| 10316 | - | -10 | -10.16 | 14.2° | - |
| 11317 | 425 | 0 | 0.00 | 14.2° | 33.5°/s |
| 12742 | 386 | 10 | -10.16 | 14.2° | 36.9°/s |
| 14127 | - | 0 | 0.00 | 14.2° | - |
| 15127 | - | 10 | -10.16 | 14.2° | - |
| 16128 | 353 | 0 | 0.00 | 14.2° | 40.3°/s |
| 17481 | 653 | 10 | 10.15 | 28.5° | 43.6°/s |
| 19134 | 607 | -10 | -10.16 | 28.5° | 46.9°/s |
| 20740 | - | 10 | 10.15 | 28.5° | - |
| 21740 | - | -10 | -10.16 | 28.5° | - |
| 22741 | 400 | 10 | 10.15 | 20.0° | 50.0°/s |
| 24141 | 566 | -10 | 10.15 | 28.5° | 50.3°/s |
| 25707 | 531 | 10 | -10.16 | 28.5° | 53.6°/s |
| 27237 | - | -10 | 10.15 | 28.5° | - |
| 28237 | - | 10 | -10.16 | 28.5° | - |
| 29237 | - | -10 | 10.15 | 28.5° | - |
| 31237 | - | 10 | -10.16 | 20.3° | - |
| 32237 | - | 10 | 10.15 | 20.0° | - |
| 33237 | - | -10 | 10.15 | 14.2° | - |

Table 1. Stimulus characteristics. Ramp stimulus invoking SP is highlighted with grey. Remaining rows describe step stimuli characteristics that invoke saccades. Value A presents the amplitude of the target's jump for saccades and magnitude of the target movements for SP. Value V represents the velocity of target's movement during for SP. Within each single time interval velocity value was a constant. Target was stationary between step and ramp signal therefore invoking fixations.

3.2 Modified Saccade Quantitative Score

The SQnS is the ratio of all detected saccade amplitudes to all saccade amplitudes encoded in the stimuli. To avoid counting corrective saccades during SP stimuli the SQnS is modified to consider saccades that directly correspond to the stimuli-saccade represented by the instantaneous jump of the target's location. To attain this goal a temporal window is introduced which considers saccades in the interval $[D_{sac_lat}, D_{sac_lat} + D_{sac_dur}]$. D_{sac_lat} is saccadic latency defined as a difference in time between the stimulus onset and the onset of the corresponding saccade. Saccadic latency for normal people is 200 ms. [1], same value is employed in our work. D_{sac_dur} is a duration of a saccade given its amplitude A_{sac} . D_{sac_dur} is measured as $2.2A_{sac} + 21$ [15]. In case of our work A_{sac} is the amplitude of the stimulus-saccade.

Ideal SQnS score, which is only achieved if the HVS perfectly executes a saccade within the temporal window and classifier accurately detects it, is 100%. In practice the SQnS value might be significantly lower because of the anticipatory and corrective saccadic behavior that falls outside of the temporal window as a result of the initial undershoots or overshoots.

3.3 Smooth Pursuit Qualitative Scores

The intuitive idea behind the Smooth Pursuit Qualitative Scores (PQIS) is to compare the proximity of the detected SP signal to the signal presented in the stimuli. Two scores are indicative of the positional (PQIS_P) and the velocity (PQIS_V) accuracy.

The PQIS_P and PQIS_V calculations are similar to the FQnS [5], i.e. for every SP point (x_s, y_s) of the presented stimuli, the check is made for the point in the eye position trace (x_e, y_e) . If such point is classified as part of SP, the Euclidean distance between these two points and the difference between their speeds are computed. Then the sum of such distances and speed differences are normalized by the amount of points compared.

$$PQIS_P = \frac{1}{N} \cdot \sum_{i=1}^N \text{pursuit_distance}_i \quad 1$$

$$PQIS_V = \frac{1}{N} \cdot \sum_{i=1}^N \text{pursuit_speed_difference}_i \quad 2$$

N is the amount of stimuli position points where stimuli pursuit state is matched with corresponding eye position sample detected as SP. $\text{pursuit_distance}_i = \sqrt{(x_s^i - x_e^i)^2 + (y_s^i - y_e^i)^2}$ and represents the distance between stimuli position and the corresponding SP point. $\text{pursuit_speed_difference}_i = |v_s^i - v_e^i|$ and represent the difference between speeds in i -th stimuli point and corresponding point in the raw eye positional sequence.

Ideal PQIS scores, which can only be achieved if HVS perfectly matches positional/velocity characteristics of the moving target and no calibration errors are present, are $PQIS_P=0^\circ$ and $PQIS_V=0^\circ/s$. In practice ideal scores might not be achieved due to calibration errors, corrective behavior, and classification inaccuracies.

3.4 Smooth Pursuit Quantitative Score

Smooth Pursuit Quantitative Score (PQnS) measures the amount of detected SP behavior given the SP behavior encoded in the stimuli. It is computed similarly to the FQnS score. Every coordinate tuple (x_s, y_s, t) of SP stimulus is compared to the corresponding tuple (x_e, y_e, t) of the eye position signal. If the recorded eye positional tuple is classified as a SP then the SP detection counter is incremented by one. The PQnS is calculated by normalizing the resulting SP behavior detection counter by the total amount of SP positional points encoded in the stimulus.

$$PQnS = 100 \cdot \frac{SP_detection_counter}{stimuli_SP_points} \quad 4$$

The computation of the ideal PQnS is performed as:

$$Ideal_PQnS = 100 \cdot \left(1 - \frac{n \cdot S_i + \sum_{j=1}^n D_{cor_sac_dur_j}}{\sum_{i=1}^n D_{stim_pur_dur_i}} \right) \quad 3$$

where n is the number of stimulus-pursuits, $D_{stim_pur_dur_i}$ is duration of the i^{th} stimulus-pursuit, S_i is pursuit's latency prior to the onset of the corrective saccade that brings the fovea to the target, and $D_{cor_sac_dur_j}$ is the expected duration of the corrective saccade. The Ideal_PQnS assumes that the HVS exhibits the SP for the duration of the target's movement immediately after the initial corrective saccade. Subsequently, accurate SP classification has to be performed for the duration of the movement. In practice ideal

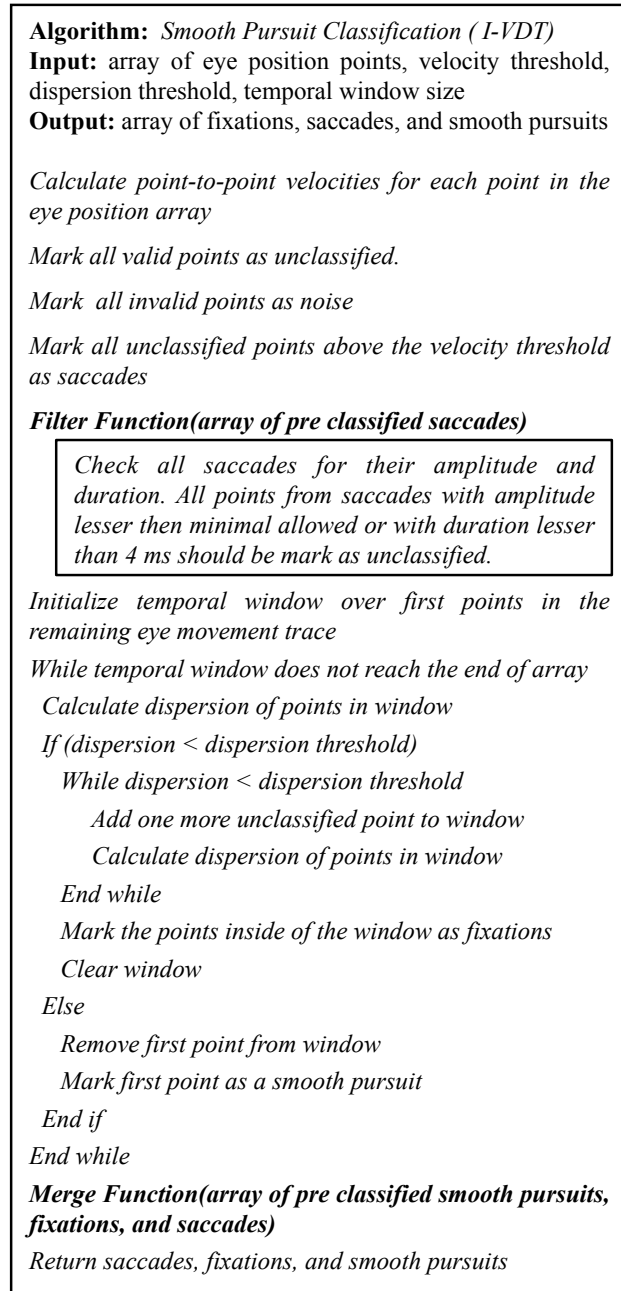


Figure 1. Pseudocode for the I-VDT algorithm

score might not be achieved due to the classification errors or additional corrective saccades occurring during the SP-stimulus.

3.5 Mismatch Scores

It is important to quantifiably assess misclassification of SP when only fixation or saccade stimulus is presented. For this purpose two mismatch scores are introduced Misclassified Fixation (Mis-Fix) and Saccade (MisSac). The similar scores that describe SP misclassified as saccades or fixations cannot be introduced because during SP stimulus HVS can actually exhibit both fixations and saccades. Therefore, it is impossible to say if fixations and

saccades during SP stimuli are detected because of the HVS behavior or misclassification caused by the classification algorithm.

To calculate MisFix, two separate quantities are calculated. $SP_fixation_points$ is the number of points in eye position trace that were classified as SP but the corresponding stimuli point for them is fixation. $total_stimuli_fixation_points$ is the total number of fixation points in stimuli

$$MisFix = 100 \cdot \frac{SP_fixation_points}{total_stimuli_fixation_points} \quad 5$$

Ideal MisFix is 0% indicating that no SP was detected during fixation stimulus. In practice ideal scores might not be achieved due to the classification errors or when the HVS stats exhibiting movements that resemble SP during fixational stimuli.

To calculate MisSac, two separate quantities are calculated. $SP_saccade_points$ is the number of points in eye position trace classified as SP. To compute the $SP_saccade_points$ value a temporal window is introduced that considers the misclassification in the interval $[D_{sac_lat}, D_{sac_lat} + D_{sac_dur}]$ that was introduced in Section 3.2. The use of such temporal window allows considering misclassification events that directly relate to the saccadic stimuli. The second value is the $total_saccades_points$. This is the number of the total eye position sample points that fit in the above mentioned interval.

$$MisSac = 100 \cdot \frac{SP_saccade_points}{total_saccades_points} \quad 6$$

Ideal MisSac is 0% indicating that no SP was detected during exhibited saccades. In practice ideal scores might not be achieved due to the classification errors or in presence of express saccades which are interspersed by miniature fixations which sometimes might resemble SP.

4 Algorithms for Smooth Pursuits Detection

I-VVT: We modify the I-VT algorithm to perform ternary classification. For the purposes of separating SP from fixations a second velocity threshold is introduced. To highlight such modification the algorithm's name is changed to the (Velocity and Velocity Threshold Identification) I-VVT. Such algorithm is capable of real-time performance, however does not provide accurate classification as shown in Section 6.

I-VMP: We call the approach proposed by San Agustin [9] and enhanced by Larsson [10] Velocity and Movement Pattern Identification (I-VMP), because it employs velocity threshold to first identify saccades similarly to I-VVT. Subsequently, it analyses the movement patterns to separate SP from fixations. The movement pattern is analyzed in a temporal window with a size T_w . In that window the direction of movement is computed by a set of angles that comprise the trajectory of raw eye positional samples in the window. The magnitude of movement is determined by the amount of angles pointing in specific direction. In our implementation we normalize the magnitude of movement to the interval [0,1] and separate the values in this interval with a threshold T_m . Values above the threshold are marked as SP and below the threshold are marked as fixations. Detailed description of the algorithm is provided elsewhere [9].

I-VDT: In this work we propose a ternary classification algorithm called Velocity and Dispersion Threshold Identification (I-VDT). It performs the initial separation of saccades similarly to the I-VVT and the I-VMP. Subsequently, it separates SP from fixations

| Test subject/Scores | S1 | S7 | S10 | Average |
|---------------------|---------------|-------------|------------|---------|
| SQnS | 37% | 32% | 60% | 43% |
| FQnS | 63% | 71% | 42% | 59% |
| PQnS | 46% | 39% | 40% | 42% |
| MisFix | 13% | 6% | 33% | 17% |
| MisSac | 11% | 0.00% | 11% | 7% |
| FQIS | 0.46° | 0.44° | 0.58° | 0.49° |
| PQIS P | 3.07° | 3.15° | 2.58° | 2.93° |
| PQIS V | 39°/s | 23°/s | 30°/s | 31°/s |
| Evaluation | medium | good | bad | |

Table 2. The scores for manually classified eye trace position for three test subjects. Their qualitative evaluation presented.

by employing a modified dispersion threshold identification method, which within a temporal window of the size T_w monitors dispersion of the points (corresponding threshold is T_d). Figure 1 presents the pseudocode. Dispersion of the points is computed in the same way as presented in [12].

5 Experimental Setup

Apparatus: The data was recorded using the EyeLink II eye tracker [22] at 1000Hz on a 21 inch CRT monitor. To ensure high accuracy of the eye movement recording a chin rest was employed. The chin rest was positioned at 70cm in front of the monitor.

Stimulus signal: Step and ramp stimulus was represented by a moving target. Target’s behavior is described in table 1 and supplied as an additional video file attached to this submission. The target was presented as a white dot with a size of approximately 1° in diameter and the center marked with a small black dot to facilitate higher targeting accuracy for HVS. Remaining screen’s background was black.

Participants: The test data consisted of a heterogeneous subject pool, age 18-25, with normal or corrected-to-normal vision. A total of 11 participants volunteered for the evaluation test. None of the participants had prior experience with eye tracking. The mean percentage of invalid data was 1.24% with maximum 7.61%. All recordings were employed during automated classification assessment (Figures 2-4). Only three recordings selected by criteria described next were employed during manual assessment.

Manual Classification: Manual classification was performed by a post-doctoral researcher to establish performance baseline and was done by visual inspection of the recorded data. Saccades were separated when the signal’s positional change was large. Fixations were separated when the signal stayed within a certain positional proximity with jitter, tremor and micro saccades present in the signal. Smooth pursuit was characterized as a signal with very low jitter and continuous directional change of the eye-gaze position. Initial corrective saccades in response to the onset of pursuit stimulus were classified as saccades

Due to considerable time necessary to classify signal manually (approximately 2.5 hours per recording) only three records were classified and were labeled as “good”, “medium”, and “bad”. “Bad” was selected because of the extremely high jitter during fixational signal. “Good” was selected due to low jitter, lack of large saccade overshoots and undershoots, and close match of SP to the stimulus position. “Medium” recording had lower jitter, however the ramp signal was not well matched by the HVS. It is

understood that the process of manual classification and categorization is subjective.

Ideal Scores: The Ideal_FQnS was computed by the eq. (4) in [5] for the stimulus characteristics described by Table 1 and was 90%. The Ideal_PQnS was computed by the eq. (1) and was 52%. Following assumptions were made for the SP latency: SP latency for stimulus-pursuits with velocity $<20^\circ/s$ is 0 ms, $<30^\circ/s$ is 230ms, $<40^\circ/s$ is 210ms, $<50^\circ/s$ is 180, and $>50^\circ/s$ is 210ms.

6 Results ternary

6.1 Manual Classification

Table 2 present the behavior scores that were computed after manual classification was done. Results indicate that the HVS performance was far from the ideal. The FQnS had the closest value to the ideal score of 84%. The SQnS was very far from the ideal value of 100%, possibly indicating that the selected performance interval of $[D_{sac\ lat}, D_{sac\ lat}+D_{sac\ dur}]$ was not realistically matched by the HVS to effectively reach target’s location. In reality undershoots, overshoots, and express saccades pushed corrective saccadic behavior out of considered boundaries. The PQnS value was far from the ideal because frequently HVS exhibited corrective saccade interspersed by fixations to follow ramp stimuli. Mismatch scores were not 0% as ideal numbers would suggest, because HVS exhibited movements that resemble SP after the ramp stimulus disappeared. All qualitative scores present reasonable values indicating relatively small positional and velocity errors between presented stimulus and position exhibited by the HVS.

In subsequent analysis of automated classification we employ the score ranges received as a result of manual classification to judge “meaningfulness” of the automated classification.

6.2 Automated Classification

Velocity threshold that separates saccades from fixations and SP was set to $70^\circ/s$ for all classification algorithms considered in this work. Such threshold was selected following the recommendations presented in [5] allowing fixing the saccade classification performance and investigating the performance of the most challenging part of the classification, i.e., separation of SP from fixations.

I-VVT: Figure 1 presents behavioral scores. The SQnS=44%

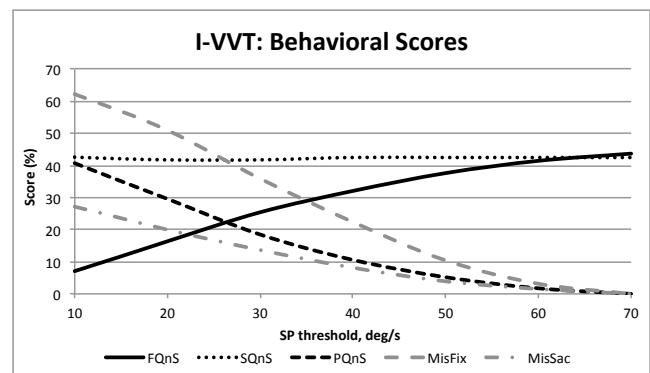


Figure 1. Performance of the I-VVT algorithm. X axis depicts SP threshold. Y axis depicts the behavioral score values

remains almost the same due to a fixed saccade threshold. The FQnS starts extremely low and increases together with the value of the SP threshold. The PQnS score starts at 41% and decreases.

Both mismatch errors start high and decrease to 0% when SP threshold reaches saccade threshold. Increase of FQnS and parallel decrease of PQnS depicts classification failure of the I-VVT.

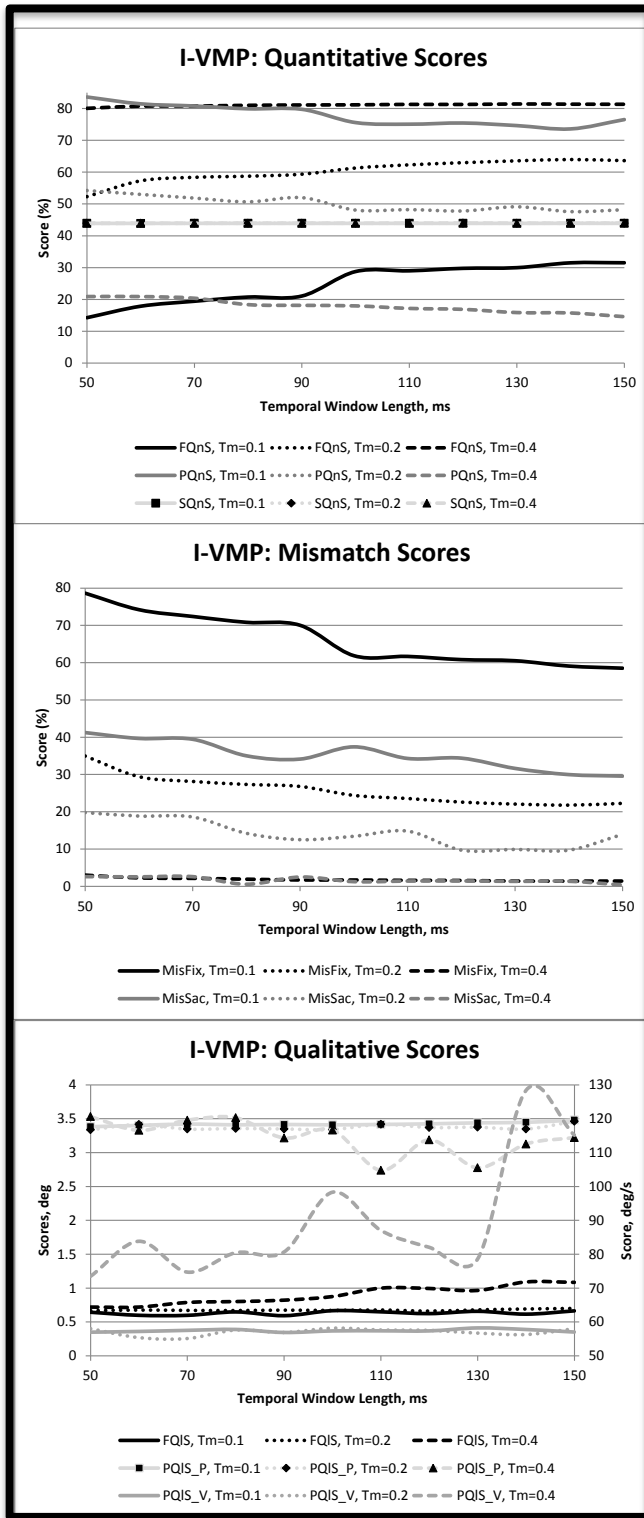


Figure 2. Performance of the I-VMP algorithm. X axis depicts SP threshold. Y axis depicts average behavioral score values.

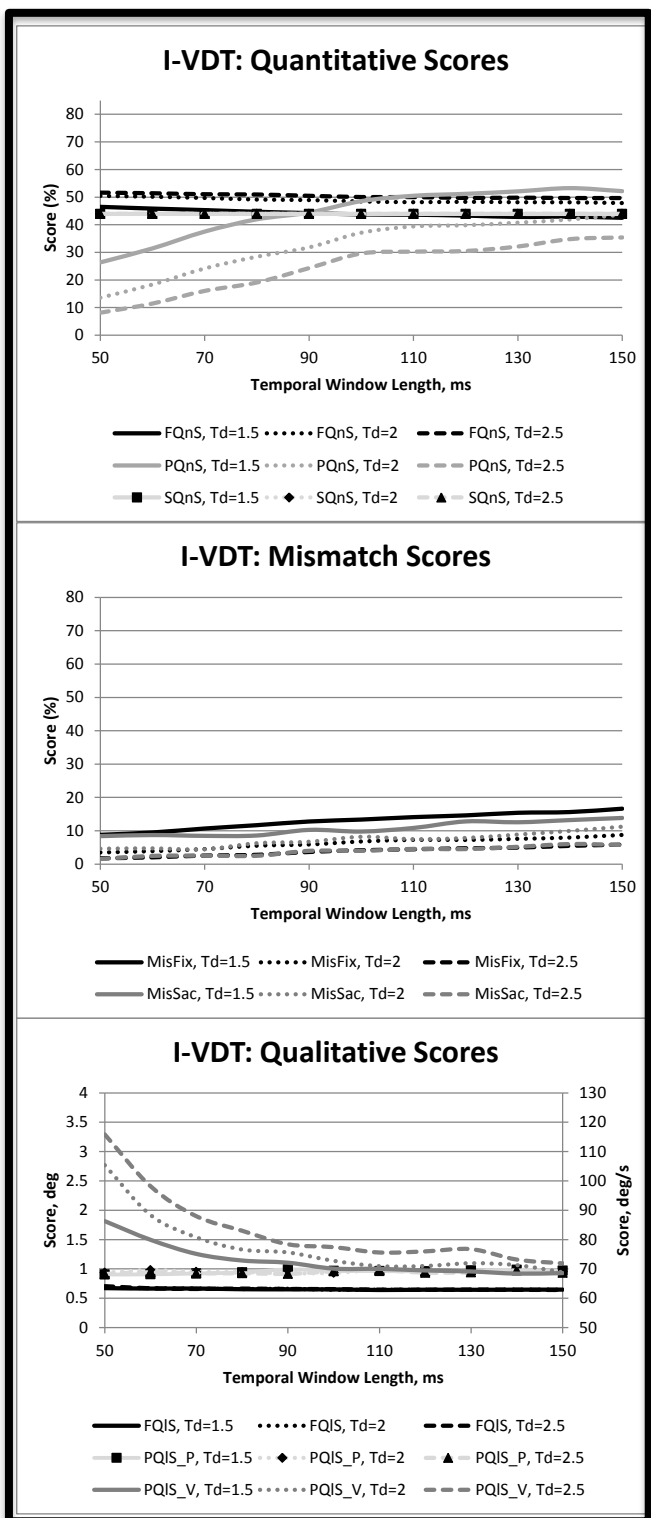


Figure 3. Performance of the I-VDT algorithm. X axis depicts SP threshold. Y axis depicts average behavioral score values.

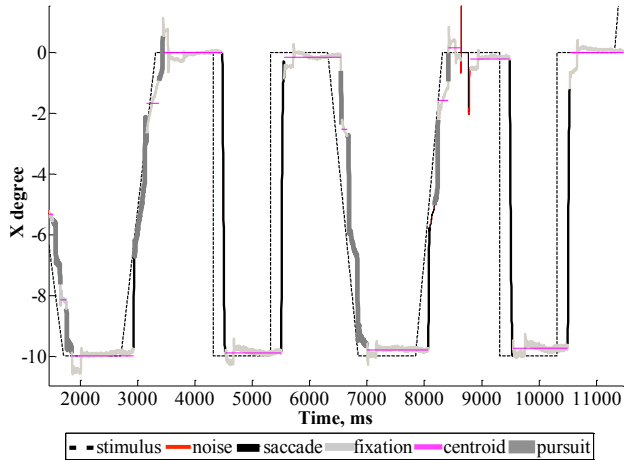


Figure 4. Results of classification performed by I-VDT for Subject 7 (S7).

which is represented by the impossibility of accurately classifying both fixations and SP at the same time. The intersection point at SP threshold of 26°/s yields $FQnS=PQnS=22\%$ is far from the values provided by manual classification. At the same time mismatch scores are too high.

I-VMP: Figure 2 presents the behavior scores. The $SQnS=44\%$ remains almost the same due to a fixed saccade threshold. The values of $FQnS$ and $PQnS$ immediately indicate that magnitude of movement threshold T_m with values of 0.1 and 0.4 does not yield acceptable classification performance indicating that the movement threshold value provides a substantial impact on classification performance. In case of $T_m=0.1$ the $FQnS$ is too low and the $PQnS$ is too high. In case of $T_m=0.4$ the $FQnS$ is too high and the $PQnS$ is too low. $T_m=0.2$ provides the most usable case, where the $FQnS$ slightly grows, when temporal window size increases, and essentially reaches the value of 63%. The $PQnS$ slightly decreases eventually reaching the value of 49% and stabilizing at that value starting temporal window threshold of $T_w=120$ ms. Obtained quantitative scores are not far from the average values depicted by the Table 2. Mismatch scores start with relatively high values but decrease with the increase of the temporal window. Both mismatch scores stabilize after temporal window reaches 120 ms. with the values, which are close to average depicted by Table 2. Starting the temporal window size of 140 ms. mismatch scores start increasing again. The $FQIS$ stays relatively stable with approximate value of 0.7° , $PQIS_V$ remains relatively stable at 58%, $PQIS_P$ fluctuates at approximately 3.5° .

I-VDT: Figure 3 presents classification performance of the I-VDT algorithm. Impact of two factors on I-VDT performance is investigated: dispersion threshold and the size of the temporal window. The increase in dispersion threshold increases $FQnS$, however slightly yielding maximum at 52%. The increase in dispersion threshold T_d significantly decreases the $PQnS$. At $T_d=1.5^\circ$ the $PQnS$ almost reaches 55%, while at $T_d=2.5^\circ$ the $PQnS$ goes down to 35%. The size of the temporal window does not impact the $FQnS$, however the growth in temporal window size produces substantial growth in $PQnS$. $PQnS$ start saturating at window sizes exceeding 110 ms. Eventually, obtained quantitative scores are not far from the average values depicted by the Table 2. Mismatch scores are higher for smaller dispersions. The growth of the temporal window makes mismatch scores grow slowly, however

without reaching 20% for either of the scores. Essentially, the mismatch scores are close to the values specified by the Table 2. Qualitative scores with the exception of the $PQnS_V$ are not affected either by the dispersion threshold or the temporal window size. The velocity error represented by the $PQnS_V$ goes down when temporal window size is increased. $PQnS_V$ value is saturated after the temporal window size reaches 110 ms. Smaller dispersion value yields smaller value.

I-VDT vs. I-VMP: From the range of thresholds that we have considered, the performance of the I-VDT is less impacted by the thresholds than the performance of the I-VMP. However, if optimum thresholds are selected for the I-VMP classification performance becomes very similar to the I-VDT. One disadvantage of the I-VMP is that the $PQnS_V$ is relatively high with a high degree of variance when compared to the I-VDT performance.

7 Discussion

7.1 Optimal Thresholds

Based on results of classifications we estimate optimal values of parameters for each used algorithm. For the **I-VVT** the optimal values of fixation threshold is 26°/s for which more or less balanced performance is achieved. However, this optimal point produces low qualitative scores and high mismatch scores when compared to the average values presented by Table 2. For the **I-VMP** the optimal value of the magnitude of movement threshold is $T_m=0.2$, with a temporal window range between the 120-140 ms. Such thresholds produce the scores that are close to the average values depicted by the Table 2. For the **I-VDT** the optimal dispersion threshold is $T_d=2^\circ$, with the temporal window of 110-150 ms. These thresholds allow to obtain scores that are close to the average scores presented by the Table 2. Classification example is depicted by Figure 5.

7.2 Limitations

The ideal behavioral scores mentioned in Sections 3 and 5 are far from what was obtained by manual classification and represented by the Table 2. This happens due to the differences in terms of what we expect from the HVS performance in the ideal scenario and what actually HVS exhibits when the stimuli that elicits three eye movement types is presented to a person. Further work is necessary to correct the performance expectations coded in the ideal score values. Such ideal score values should be useful in selecting the thresholds that provide meaningful classification performance.

A very specific hardware and step ramp stimulus was employed in this work. A chin rest was employed for additional stability of the recorded data. Future research is necessary to provide more comprehensive performance picture of ternary eye movement classification algorithms that employs different hardware, allows freedom of head movement, and contains different stimuli characteristics. However, we expect that behavioral scores performance will remain similar to our findings here.

7.3 HVS performance

During manual inspection of the recorded data we have noticed substantial variability of HVS performance between and within subjects. SP latency, the size of the corrective saccades and the quality of target's tracking vary substantially. Very frequently during ramp stimulus the HVS exhibits a sequence of corrective

saccades interspersed by fixations, rather than tracking the target smoothly. Closer to the end of the recording, when a subject has experienced a variety of SP stimuli, the HVS started exhibiting during fixational stimulus occasional movements with characteristics resembling the SP. MisFix errors presented in Table 2 highlight this peculiarity. Such HVS performance further complicates ternary classification and necessitates very careful construction of the ideal behavioral scores.

8 Conclusions & Future Work

This paper considered and introduced methods for reliable automated ternary classification that consists of three eye movement types: fixations, saccades, and smooth pursuit. This task is extremely challenging due to the substantial variability of Human Visual System performance between and within subjects, difficulties in separation of fixations from smooth pursuit, and substantial noisiness of the eye tracking data.

We have extended the set of behavior scores originally introduced by Komogortsev and colleagues [5] with a purpose of assessing the meaningfulness of ternary classification.

We should point out that the performance of the HVS was different from the expectations that were encoded in the ideal behavior scores. This difference in performance contrasts with the scenario when only fixations and saccades are present in the data, in which case the ideal behavior scores and HVS performance are very similar, therefore allowing to select meaningful thresholds based on the ideal score values [5]. As a result further research is necessary to define what realistically can be expected from the HVS if fixations, saccades, and smooth pursuit eye movement types are present in the eye positional data. In case of this work, manual classification of the recorded data was performed and scored to serve as a performance baseline for the assessment of the automated classification algorithms.

Our findings indicate that a simple extension of the popular velocity threshold method (I-VT) algorithm with an idea of separating fixations from smooth pursuit with an auxiliary velocity threshold will not provide meaningful ternary classification. Two additional algorithms were considered Velocity Movement Pattern Identification (I-VMP) as introduced by San Agustin [9] and Larsson [10] and the algorithm that we have developed in this work Velocity Dispersion Threshold Identification (I-VDT). Both algorithms when driven by the optimal thresholds were able to provide classification results that were close to the results obtained via manual classification. However, within considered threshold intervals the I-VDT had smaller performance variability and dependence on the thresholds than the I-VMP possibly indicating higher practical usefulness. Misclassification errors were also slightly smaller for the proposed I-VDT algorithm. Classification speed was very fast for both algorithms and only was limited by the size of the temporal window.

Future work will adjust expectations encoded in the ideal behavioral scores and will seek stimulus presentation that can be effectively presented as a part of the eye tracker calibration procedure to be able to suggest optimal threshold selection for any algorithm that performs ternary classification. Such automated threshold selection should be particularly useful for eye tracking practitioners that would be able to use suggested thresholds for a variety of stimuli recorded immediately after the calibration procedure.

9 Acknowledgement

This work was partially funded by the National Institute of Standards grant #60NANB10D213.

References

- [1] R. J. Leigh and D. S. Zee, *The Neurology of Eye Movements*: Oxford University Press, 2006.
- [2] H. Istance, A. Hyrskykari, L. Immonen, S. Mansikkamaa, and S. Vickers, "Designing gaze gestures for gaming: an investigation of performance," presented at the Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications, Austin, Texas, 2010.
- [3] E. Isotalo, A. Heikki, and P. Ilmari, "Oculomotor findings mimicking a cerebellar disorder and postural control in severe Meniere's disease," *Auris Nasus Larynx*, vol. 36, pp. 36-41, 2009.
- [4] F. Shic, K. Chawarska, and B. Scassellati, "The Amorphous Fixation Measure Revisited: with Applications to Autism," in *30th Annual Meeting of the Cognitive Science Society (CogSci 2008)*, 2008, pp. 1-6.
- [5] O. V. Komogortsev, D. V. Gobert, S. Jayarathna, D. Koh, and S. Gowda, "Standardization of Automated Analyses of Oculomotor Fixation and Saccadic Behaviors," *IEEE Transactions on Biomedical Engineering*, vol. 57, pp. 2635-2645, 2010.
- [6] M. Nyström and K. Holmqvist, "An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data," *Behavior Research Methods*, vol. 42, pp. 188-204, 2010.
- [7] D. W. Hansen and J. Qiang, "In the Eye of the Beholder: A Survey of Models for Eyes and Gaze," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 478-500, 2010.
- [8] S. Lebedev, P. Van Gelder, and W. H. Tsui, "Square-root relations between main saccadic parameters," *Investigative Ophthalmology & Visual Science*, vol. 37, pp. 2750-8, 1996.
- [9] J. S. Agustin, "Off-the-Shelf Gaze Interaction," PhD Thesis, The IT University of Copenhagen, 2009.
- [10] L. Larsson, "Event Detection in Eye-Tracking Data," Master Thesis, Department of Electrical and Information Technology, Lund University, Lund, Sweden, 2010.
- [11] C. H. Meyer, A. G. Lasker, and D. A. Robinson, "The Upper Limit of Human Smooth Pursuit Velocity," *Vision Research*, vol. 25, pp. 561-563, 1985.
- [12] D. D. Salvucci and J. H. Goldberg, "Identifying fixations and saccades in eye tracking protocols," in *Eye Tracking Research and Applications Symposium*, New York, 2000, pp. 71-78.
- [13] J. R. Tole and L. R. Young, "Digital Filters for Saccade and Fixation Detection," in *Eye Movements: Cognition and Visual Perception*, D. Fisher, R. Monty, and J. W. Senders, Eds., ed Hillsdale, NJ: Lawrence Erlbaum Associates, 1981, pp. 185-199.
- [14] D. A. Robinson, "The Mechanics of Human Smooth Pursuit Eye Movement," *Journal of Physiology*, vol. 180, pp. 569-591, 1965.
- [15] R. H. S. Carpenter, *Movements of the Eyes*. London: Pion, 1977.
- [16] D. I. Braun, N. Mennie, C. Rasche, A. C. Schutz, M. J. Hawken, and K. R. Gegenfurtner, "Smooth Pursuit Eye

- Movements to Isoluminant Targets," *Journal of Neurophysiology*, vol. 100, pp. 1287-1300, 2008.
- [17] J.-J. O. d. Xivry, M. Missal, and P. Lefèvre, "Smooth pursuit performance during target blanking does not influence the triggering of predictive saccades," *Journal of Vision*, vol. 9, pp. 1–16, 2009.
- [18] T. Shibata, H. Tabata, S. Schaal, and M. Kawato, "A model of smooth pursuit in primates based on learning the target dynamics," *Neural Networks*, vol. 18, pp. 213-224, 2005.
- [19] A. Tavassoli and D. L. Ringach, "Dynamics of Smooth Pursuit Maintenance," *Journal of Neurophysiology*, vol. 102, pp. 110–118, 2009.
- [20] R. Lencer and P. Trillenber, "Neurophysiology and Neuroanatomy of Smooth Pursuit in Humans," *Brain and Cognition*, vol. 68, pp. 219–228, 2008.
- [21] A. T. Bahill, M. J. Iandolo, and B. T. Troost, "Smooth Pursuit Eye Movements in Response to Unpredictable Target Waveforms," *Vision Research*, vol. 20, pp. 923-931, 1980.
- [22] EyeLink. (2010). *EyeLink II*. Available: http://www.sr-research.com/EL_1000.html